Probing the correlations in composite signals

S. Sargsyan

Center for Cosmology and Astrophysics, Alikhanian National Laboratory and Yerevan State University, Yerevan, Armenia

E-mail: `seda@yerphi.am`

Abstract. The technique of degree of randomness is used to model the correlations in sequences containing various subsignals and noise. Kolmogorov stochasticity parameter enables to quantify the randomness in number sequences and hence appears as an efficient tool to distinguish the signals. Numerical experiments for a broad class of composite signals of regular and random properties enable to obtain the qualitative and quantitative criteria for the behavior of the descriptor depending on the input parameters typical to astrophysical signals.

1. The method

The correlations functions and the power spectra are common and efficient tools for the study of correlations in the signals. Astrophysical signals typically are superposition of various subsignals, regular and random, by features comparable to each other and of weaker ones, i.e. perturbations or the noise. The procedure of analysing of the needed signal or signals, their separation from the noise is a common problem while dealing with observations and measurements.

The Kolmogorov stochasticity parameter technique enables to quantify the randomness of sequences of number theory or dynamical systems [1, 2, 3, 4, 5].

The technique of the degree of randomness has been applied to the Cosmic Microwave Background (CMB) temperature sky maps and to the X-ray flux data of the clusters of galaxies. The former data were those obtained by the Wilkinson Microwave Anisotropy Probe (WMAP) during 7-year observations [6, 7], while the X-ray data were obtained by XMM-Newton satellite providing a particularly accurate and complete sky survey (see [8, 9]). In the case of CMB, the Kolmogorov function enabled to separate signals of different origin, e.g. the Galactic and non-Galactic ones, and to detect point sources in the CMB maps [10] (see Fig.1). Concerning the X-ray clusters, it was shown that their X-ray images do possess correlation in the pixelized flux data peculiar to the gravitational potential of the galaxy clusters [11]. This technique resembles the methods of the dynamical systems applied to nonlinear problems (e.g. [12]).

A crucial step in these studies is the modeling and analysis of generated systems, which enables to reveal the behavior of the stochasticity parameter in the case of a given signal and then to consider the application of this technique for real signals [13]. Below we represent the results of numerical experiments for a broad class of signals.

Kolmogorov stochasticity parameter is introduced for a sequence $\{X_1, X_2, \ldots, X_n\}$ of real random variable X sorted in growing order $X_1 \leq X_2 \leq \ldots \leq X_n$. Then the theoretical distribution function is [1, 2]

$$F(X) = n \cdot (probability\ of\ the\ event\ x \leq X). \tag{1}$$

the stochasticity parameter is defined as

$$\lambda_n = \sqrt{n} \, \sup_x |F_n(x) - F(x)| \tag{2}$$

where the empirical distribution function is

$$F_n(X) = (number\ of\ the\ elements\ x_i\ which\ are\ less\ than\ X),$$

and

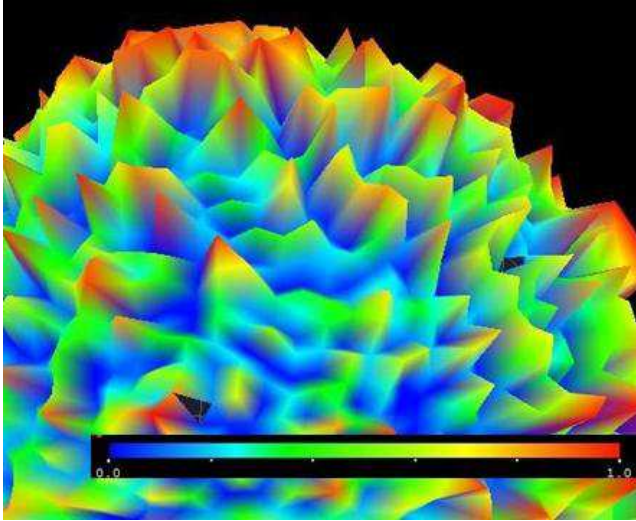$$F_n(X) = \begin{cases} 0, & X < x_1 \\ k/n, & x_k \le X < x_{k+1} \\ 1, & x_n \le X. \end{cases} \tag{3}$$

Then for the limit

$$\lim_{n\to\infty} P\{\lambda_n \le \lambda\} = \Phi(\lambda) \,, \tag{4}$$

where

$$\Phi(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k \, e^{-2k^2\lambda^2}, \, \Phi(0) = 0, \, \lambda > 0 \,, \tag{5}$$

exists at uniform convergence and independent on F.

Figure 1. The Kolmogorov's function Φ for the portion of the 7-year temperature CMB map obtained by WMAP.

2. Random-regular sequences

We consider a broad class of sequences, i.e. those composed of random x_n and regular $y_n = \frac{an \pmod b}{b}$ (a, b are prime numbers) sub-sequences within $(0, 1)$

$$z_n = \alpha x_n + (1 - \alpha)y_n. \tag{6}$$

The parameter α varies within [0,1] defining random sequences at 1 and regular ones at 0, so that by mutually fixing the pair a, b we get new regular sequences.

For z_n we have

$$F(X) = \begin{cases} 0, & X \leq 0 \\ \frac{X^2}{2\alpha(1-\alpha)}, & 0 < X \leq \alpha \\ \frac{2\alpha X - \alpha^2}{2\alpha(1-\alpha)}, & \alpha < X \leq 1 - \alpha \\ 1 - \frac{(1-X)^2}{2\alpha(1-\alpha)}, & 1 - \alpha < X \leq 1 \\ 1, & X > 1. \end{cases} \qquad (7)$$

Figure 2 shows the results of the numerical experiments for 100 sequences, each sequence containing 10000 elements. Each sequence is divided into 50 subsequences, i.e. m runs through values $1, ..., 50$, and for each of them the parameter $\Phi(\lambda_n)_m$ is calculated and then the empirical distribution function $G(\Phi)_m$ of these numbers is obtained. When the original sequences are random, this distribution have to be uniform according to Kolmogorov's theorem. To test that, χ^2 for the functions $G(\Phi)_m$ and $G_0(\Phi) = \Phi$ have been calculated, i.e. one parameter χ^2 is calculated for each of the 100×101 sequences. For 100 χ^2 values per each value of α, we obtained the mean and error values for χ^2, i.e. for each pair a, b we have a plot of the dependence of χ^2 on α.
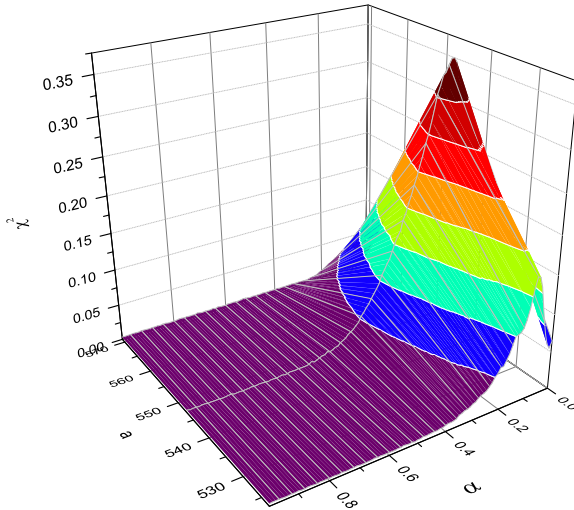

Figure 2. The 3D χ^2 for the Kolmogorov's function for the sequence z_n vs α and the parameter a.

3. Parameters of the regular sequences

At certain values of the parameter a for different values of b the dependences in Fig.2 are monotonic, while for others they do have maxima. To study this effect, we introduce a parameter Δ which is the difference of two values in those plots: maximal value of χ^2 and minimal value in the range $\alpha \in (0, \alpha_{max})$, if α_{max} is the position of the maximal value. Obviously, Δ is zero when the dependence is monotonic and no extrema do exist. Then we calculate Δ for fixed b and for each value of primary $a = 2, ..., b$.

The remarkable feature of the results is the strict mirror symmetry in Fig.3 in the dependence of Δ vs a, although no periodicity is found by Fourier analysis.

The mirror symmetric plots can be hence subjects of particular study, e.g. in two versions: first, of the distribution of Δ and, second, the spacing between non-zero Δ and their distribution.
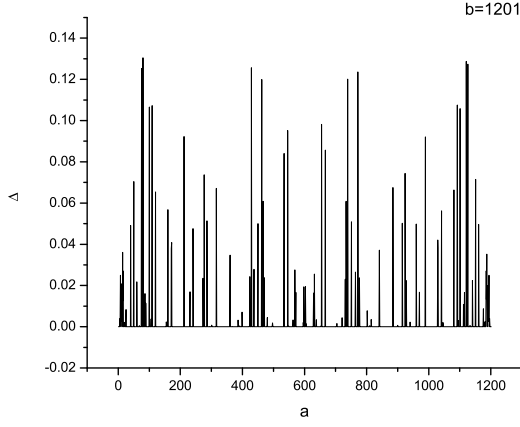
Figure 3. The dependence of Δ vs the parameter a.

The null values of Δ are skipped and also - since due to the mirror symmetry each Δ has its pair - only one of each pair is taken into account.

The results we give in Fig.4 where the number of non-zero Δs from Fig.3 is given in growing order. The number of non-zero Δs appear to be proportional to b.
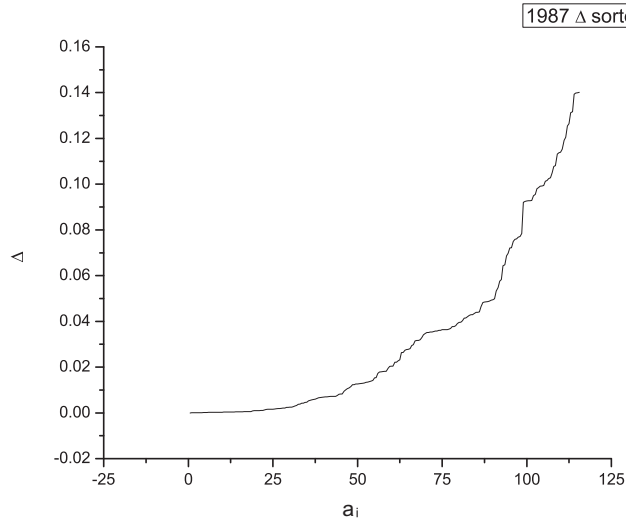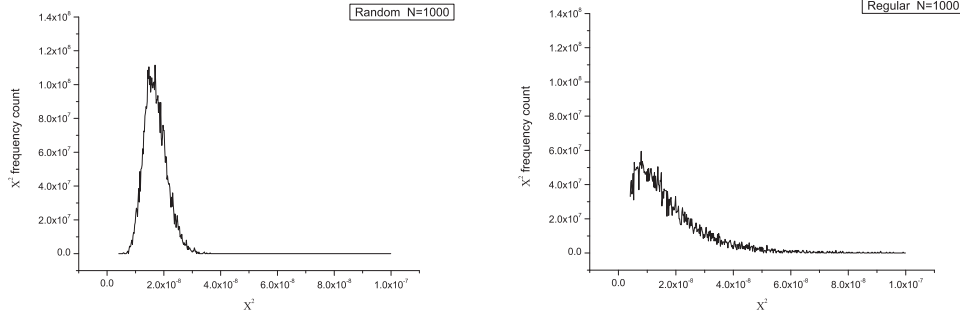

Figure 4. Sorted amplitudes of Δs from Fig.3.

4. Sum of fluctuations: large N limit

The next problem we consider is the properties of the signal being a sum of random and regular fluctuations, each of sequences of 10000 elements and of the same standard deviation.

The regular sequences have been chosen as

$$a_i = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} Compact(x_i^k, -1, 1), \tag{8}$$

Figure 5. χ^2 frequency for random and regular sequences vs the Gaussian distribution.
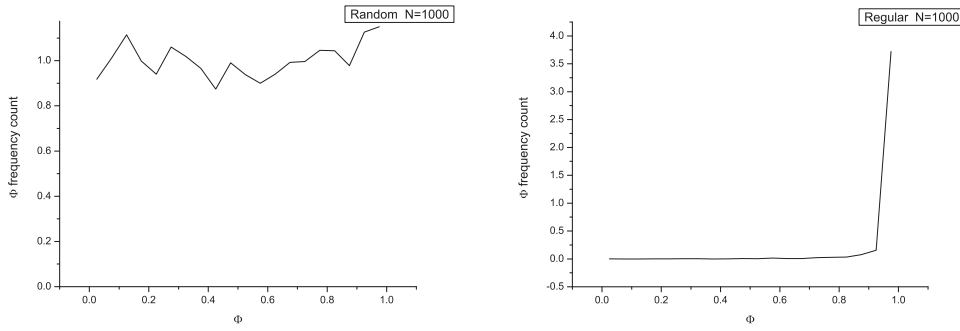
Figure 6. Kolmogorov function Φ for the random and regular sequences in Fig.5.

where

$$x_i^k = i/k,$$

x^k is compactified arithmetical sequence within the interval $(-1, 1)$, with step $1/k$.

The random sequences are given by

$$b_i = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} Random(-1, 1).$$

At large number of sequences each new sequence y_n is taken as the continuation of the former arithmetical progression.

Here

$$Compact(x, p, q) = q + x mod(p - q)$$

indicate multiples of $(p - q)$ from x having the value within the range $(p, q), p < q$.

The results for random and regular sequences, 10000 each, are given in Fig.5 for χ^2, when the number of the fluctuations vary from $N = 1000$ to 100000. The χ^2 shows that for both, random and regular sequences, we deal with a Gaussian limiting distribution, in accordance with the Central Limit theorem which states that for large enough values of N, both sequences a_i and b_i tend to Gaussian sequences with the same σ and μ independent on N. So, although differences are seen in the Gaussians, namely, the standard deviations are larger for the regular case, the χ^2 are similar.

For the Kolmogorov function Φ the situation is rather different. When the Gaussians do appear both for random and regular sequences (as expected), the behaviors of Φ is different and enables to separate them, as shown in Fig.6. Namely, it is close to a homogeneous function for random sequences and $\Phi = 1$ for regular ones. Kolmogorov's function therefore enables to

distinguish the superposition of random and regular sequences, even though both are tending to Gaussians.

Finally, we have probed the dependence of the results on the length of the sequences: the dependence on the number of the fluctuations within $1000 - 100000$ is rather weak, χ^2 varying around $10^{-8} - 10^{-9}$. This confirms the universality of the obtained behavior of Φ for both random and regular fluctuations.

5. Results

The performed analysis revealed the behavior of the Kolmogorov distribution vs the properties of the generated signals. To describe datasets which contain both regular and stochastic components, we considered sequences scaled by a single parameter α, indicating the ratio of those components.

Quantitative and qualitative criteria have been obtained for the Kolmogorov distribution at numerical experiments for broad class of random and regular sequences depending on α parameter.

a) The existence of the *critical value* for α has been shown, when the monotonic decay of the frequency count of the Kolmogorov distribution is transformed to a function with an extremum.

b) The dependence of scalings and spacings vs that parameter shows mirror properties both in the amplitude and distribution of the frequency counts of the function Φ.

c) The behavior of the randomness of a signal composed of N subsignals at large N limit has been studied, where the Kolmogorov function acts as an informative descriptor. Particularly, the descriptor at large N enables to distinguish the initial set of the fluctuations, even when the superposition both of random and regular subsignals is not informative since tends to a Gaussian in accordance to the Central Limit theorem.

The studied properties are typical, for example, for astrophysical datasets, when the sought signals are superposed with regular and random fluctuations of various origin, and hence the behaviors revealed at the numerical experiments due to the universality of the technique will enable its informative application to real data.

[1] Kolmogorov A.N. 1933 *G.Ist.Ital.Attuari,* **4** 83
[2] Arnold V.I. 2008 *Nonlinearity* **21** T109
[3] Arnold V.I. 2008 *Uspekhi Mat. Nauk* **63** 5
[4] Arnold V.I. 2009 *Trans. Mosc. Math. Soc.* **70** 31
[5] Arnold V.I. 2009 *Funct. Anal. Other Math.* **2** 139
[6] Komatsu E., Dunkley J. *et al.* 2009 *ApJS* **180** 330
[7] Jarosik N., Bennett C.L. *et al.* 2011 *ApJS* **192** 14
[8] Viana P.T.P., da Silva A. *et al.* 2011 *arXiv:1109.1828*
[9] Suhada R., Song J., *et al.* 2011 *arXiv:1111.0141*
[10] Gurzadyan V.G., Allahverdyan A.E. *et al.* 2009 *Astron. & Astrophys.* **497** 343
[11] Gurzadyan V.G., Durret F. *et al.* 2011 *Europhys.Lett.* **95** 69001
[12] Gurzadyan V.G., Pfenniger D., (Eds.) 1994 *Ergodic Concepts in Stellar Dynamics,* Springer-Verlag.
[13] Gurzadyan V.G., Ghahramanyan T., Sargsyan S. 2011 *Europhys.Lett.* **95** 19001